

RAW SEQUENCE LISTING  
PATENT APPLICATION US/09/189,028DATE: 08/25/1999  
TIME: 16:39:31

INPUT SET: S33065.raw

This Raw Listing contains the General  
Information Section and up to the first 5 pages.

#4

ENTERED

## SEQUENCE LISTING

1  
2  
3 (1) General Information:  
4  
5 (i) APPLICANT: Rasmussen, Grethe  
6 Mikkelsen, Jan Moller  
7 Schulein, Martin  
8 Patkar, Shankant A.  
9 Hagen, Fred  
10  
11 (ii) TITLE OF INVENTION: A Cellulase Preparation Comprising an  
12 Endoglucanase Enzyme  
13  
14 (iii) NUMBER OF SEQUENCES: 33  
15  
16 (iv) CORRESPONDENCE ADDRESS:  
17 (A) ADDRESSEE: Novo Nordisk of North America, Inc.  
18 (B) STREET: 405 Lexington Avenue, 64th Floor  
19 (C) CITY: New York  
20 (D) STATE: New York  
21 (E) COUNTRY: United States of America  
22 (F) ZIP: 10174-6401  
23  
24 (v) COMPUTER READABLE FORM:  
25 (A) MEDIUM TYPE: Floppy disk  
26 (B) COMPUTER: IBM PC compatible  
27 (C) OPERATING SYSTEM: PC-DOS/MS-DOS  
28 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30  
29  
30 (vi) CURRENT APPLICATION DATA:  
31 (A) APPLICATION NUMBER: 09/189,028  
32 (B) FILING DATE:  
33 (C) CLASSIFICATION:  
34  
35 (vii) PRIOR APPLICATION DATA:  
36 (A) APPLICATION NUMBER: US 08/389,423  
37 (B) FILING DATE: 14-FEB-1995  
38  
39 (viii) ATTORNEY/AGENT INFORMATION:  
40 (A) NAME: Lambiris, Elias J.  
41 (B) REGISTRATION NUMBER: 33,728  
42 (C) REFERENCE/DOCKET NUMBER: 3469.214-US  
43  
44 (ix) TELECOMMUNICATION INFORMATION:  
45 (A) TELEPHONE: 212-867-0123  
46 (B) TELEFAX: 212-878-9655

RAW SEQUENCE LISTING  
PATENT APPLICATION US/09/189,028DATE: 08/25/1999  
TIME: 16:39:31

INPUT SET: S33065.raw

47  
48  
49 (2) INFORMATION FOR SEQ ID NO:1:  
50  
51 (i) SEQUENCE CHARACTERISTICS:  
52 (A) LENGTH: 1060 base pairs  
53 (B) TYPE: nucleic acid  
54 (C) STRANDEDNESS: single  
55 (D) TOPOLOGY: linear  
56  
57 (ii) MOLECULE TYPE: cDNA  
58  
59 (iii) HYPOTHETICAL: NO  
60  
61 (vi) ORIGINAL SOURCE:  
62 (A) ORGANISM: Humicola insolens  
63 (B) STRAIN: DSM 1800  
64  
65 (ix) FEATURE:  
66 (A) NAME/KEY: mat\_peptide  
67 (B) LOCATION: 73..924  
68  
69 (ix) FEATURE:  
70 (A) NAME/KEY: sig\_peptide  
71 (B) LOCATION: 10..72  
72  
73 (ix) FEATURE:  
74 (A) NAME/KEY: CDS  
75 (B) LOCATION: 10..924  
76  
77 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:  
78  
79 GGATCCAAG ATG CGT TCC TCC CCC CTC CTC CCG TCC GCC GTT GTG GCC 48  
80 Met Arg Ser Ser Pro Leu Leu Pro Ser Ala Val Val Ala  
81 -21 -20 -15 -10  
82  
83 GCC CTG CCG GTG TTG GCC CTT GCC GCT GAT GGC AGG TCC ACC CGC TAC 96  
84 Ala Leu Pro Val Leu Ala Leu Ala Ala Asp Gly Arg Ser Thr Arg Tyr  
85 -5 1 5  
86  
87 TGG GAC TGC TGC AAG CCT TCG TGC GGC TGG GCC AAG AAG GCT CCC GTG 144  
88 Trp Asp Cys Cys Lys Pro Ser Cys Gly Trp Ala Lys Lys Ala Pro Val  
89 10 15 20  
90  
91 AAC CAG CCT GTC TTT TCC TGC AAC GCC AAC TTC CAG CGT ATC ACG GAC 192  
92 Asn Gln Pro Val Phe Ser Cys Asn Ala Asn Phe Gln Arg Ile Thr Asp  
93 25 30 35 40  
94  
95 TTC GAC GCC AAG TCC GGC TGC GAG CCG GGC GGT GTC GCC TAC TCG TGC 240  
96 Phe Asp Ala Lys Ser Gly Cys Glu Pro Gly Gly Val Ala Tyr Ser Cys  
97 45 50 55  
98  
99 GCC GAC CAG ACC CCA TGG GCT GTG AAC GAC GAC TTC GCG CTC GGT TTT 288

# RAW SEQUENCE LISTING PATENT APPLICATION US/09/189,028

DATE: 08/25/1999  
TIME: 16:39:32

INPUT SET: S33065.raw

100	Ala	Asp	Gln	Thr	Pro	Trp	Ala	Val	Asn	Asp	Asp	Phe	Ala	Leu	Gly	Phe	
101				60					65					70			
102																	
103	GCT	GCC	ACC	TCT	ATT	GCC	GGC	AGC	AAT	GAG	GCG	GGC	TGG	TGC	TGC	GCC	336
104	Ala	Ala	Thr	Ser	Ile	Ala	Gly	Ser	Asn	Glu	Ala	Gly	Trp	Cys	Cys	Ala	
105			75					80					85				
106																	
107	TGC	TAC	GAG	CTC	ACC	TTC	ACA	TCC	GGT	CCT	GTT	GCT	GGC	AAG	AAG	ATG	384
108	Cys	Tyr	Glu	Leu	Thr	Phe	Thr	Ser	Gly	Pro	Val	Ala	Gly	Lys	Lys	Met	
109		90					95					100					
110																	
111	GTC	GTC	CAG	TCC	ACC	AGC	ACT	GGC	GGT	GAT	CTT	GGC	AGC	AAC	CAC	TTC	432
112	Val	Val	Gln	Ser	Thr	Ser	Thr	Gly	Gly	Asp	Leu	Gly	Ser	Asn	His	Phe	
113	105					110					115					120	
114																	
115	GAT	CTC	AAC	ATC	CCC	GGC	GGC	GGC	GTC	GGC	ATC	TTC	GAC	GGA	TGC	ACT	480
116	Asp	Leu	Asn	Ile	Pro	Gly	Gly	Gly	Val	Gly	Ile	Phe	Asp	Gly	Cys	Thr	
117					125					130					135		
118																	
119	CCC	CAG	TTC	GGC	GGT	CTG	CCC	GGC	CAG	CGC	TAC	GGC	GGC	ATC	TCG	TCC	528
120	Pro	Gln	Phe	Gly	Gly	Leu	Pro	Gly	Gln	Arg	Tyr	Gly	Gly	Ile	Ser	Ser	
121				140					145					150			
122																	
123	CGC	AAC	GAG	TGC	GAT	CGG	TTC	CCC	GAC	GCC	CTC	AAG	CCC	GGC	TGC	TAC	576
124	Arg	Asn	Glu	Cys	Asp	Arg	Phe	Pro	Asp	Ala	Leu	Lys	Pro	Gly	Cys	Tyr	
125			155					160					165				
126																	
127	TGG	CGC	TTC	GAC	TGG	TTC	AAG	AAC	GCC	GAC	AAT	CCG	AGC	TTC	AGC	TTC	624
128	Trp	Arg	Phe	Asp	Trp	Phe	Lys	Asn	Ala	Asp	Asn	Pro	Ser	Phe	Ser	Phe	
129		170					175					180					
130																	
131	CGT	CAG	GTC	CAG	TGC	CCA	GCC	GAG	CTC	GTC	GCT	CGC	ACC	GGA	TGC	CGC	672
132	Arg	Gln	Val	Gln	Cys	Pro	Ala	Glu	Leu	Val	Ala	Arg	Thr	Gly	Cys	Arg	
133	185					190					195					200	
134																	
135	CGC	AAC	GAC	GAC	GGC	AAC	TTC	CCT	GCC	GTC	CAG	ATC	CCC	TCC	AGC	AGC	720
136	Arg	Asn	Asp	Asp	Gly	Asn	Phe	Pro	Ala	Val	Gln	Ile	Pro	Ser	Ser	Ser	
137					205					210					215		
138																	
139	ACC	AGC	TCT	CCG	GTC	AAC	CAG	CCT	ACC	AGC	ACC	AGC	ACC	ACG	TCC	ACC	768
140	Thr	Ser	Ser	Pro	Val	Asn	Gln	Pro	Thr	Ser	Thr	Ser	Thr	Thr	Ser	Thr	
141				220					225					230			
142																	
143	TCC	ACC	ACC	TCG	AGC	CCG	CCA	GTC	CAG	CCT	ACG	ACT	CCC	AGC	GGC	TGC	816
144	Ser	Thr	Thr	Ser	Ser	Pro	Pro	Val	Gln	Pro	Thr	Thr	Pro	Ser	Gly	Cys	
145			235					240					245				
146																	
147	ACT	GCT	GAG	AGG	TGG	GCT	CAG	TGC	GGC	GGC	AAT	GGC	TGG	AGC	GGC	TGC	864
148	Thr	Ala	Glu	Arg	Trp	Ala	Gln	Cys	Gly	Gly	Asn	Gly	Trp	Ser	Gly	Cys	
149		250					255					260					
150																	
151	ACC	ACC	TGC	GTC	GCT	GGC	AGC	ACT	TGC	ACG	AAG	ATT	AAT	GAC	TGG	TAC	912
152	Thr	Thr	Cys	Val	Ala	Gly	Ser	Thr	Cys	Thr	Lys	Ile	Asn	Asp	Trp	Tyr	

# RAW SEQUENCE LISTING

## PATENT APPLICATION US/09/189,028

DATE: 08/25/1999  
TIME: 16:39:32

INPUT SET: S33065.raw

```

153      265      270      275      280
154
155 CAT CAG TGC CTG TAGACGCAGG GCAGCTTGAG GGCCTTACTG GTGGCCGCAA      964
156 His Gln Cys Leu
157
158
159 CGAAATGACA CTCCCAATCA CTGTATTAGT TCTTGACAT AATTTCGTCA TCCCTCCAGG      1024
160
161 GATTGTCACA TAAATGCAAT GAGGAACAAT GAGTAC      1060
162
163
164 (2) INFORMATION FOR SEQ ID NO:2:
165
166 (i) SEQUENCE CHARACTERISTICS:
167 (A) LENGTH: 305 amino acids
168 (B) TYPE: amino acid
169 (D) TOPOLOGY: linear
170
171 (ii) MOLECULE TYPE: protein
172
173 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:
174
175 Met Arg Ser Ser Pro Leu Leu Pro Ser Ala Val Val Ala Ala Leu Pro
176 -21 -20      -15      -10
177
178 Val Leu Ala Leu Ala Ala Asp Gly Arg Ser Thr Arg Tyr Trp Asp Cys
179 -5      1      5      10
180
181 Cys Lys Pro Ser Cys Gly Trp Ala Lys Lys Ala Pro Val Asn Gln Pro
182      15      20      25
183
184 Val Phe Ser Cys Asn Ala Asn Phe Gln Arg Ile Thr Asp Phe Asp Ala
185      30      35      40
186
187 Lys Ser Gly Cys Glu Pro Gly Gly Val Ala Tyr Ser Cys Ala Asp Gln
188      45      50      55
189
190 Thr Pro Trp Ala Val Asn Asp Asp Phe Ala Leu Gly Phe Ala Ala Thr
191      60      65      70      75
192
193 Ser Ile Ala Gly Ser Asn Glu Ala Gly Trp Cys Cys Ala Cys Tyr Glu
194      80      85      90
195
196 Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Lys Met Val Val Gln
197      95      100      105
198
199 Ser Thr Ser Thr Gly Gly Asp Leu Gly Ser Asn His Phe Asp Leu Asn
200      110      115      120
201
202 Ile Pro Gly Gly Gly Val Gly Ile Phe Asp Gly Cys Thr Pro Gln Phe
203      125      130      135
204
205 Gly Gly Leu Pro Gly Gln Arg Tyr Gly Gly Ile Ser Ser Arg Asn Glu

```

# RAW SEQUENCE LISTING PATENT APPLICATION US/09/189,028

DATE: 08/25/1999  
TIME: 16:39:33

INPUT SET: S33065.raw

206	140	145	150	155
207				
208	Cys Asp Arg Phe Pro Asp Ala Leu Lys Pro Gly Cys Tyr Trp Arg Phe			
209		160	165	170
210				
211	Asp Trp Phe Lys Asn Ala Asp Asn Pro Ser Phe Ser Phe Arg Gln Val			
212		175	180	185
213				
214	Gln Cys Pro Ala Glu Leu Val Ala Arg Thr Gly Cys Arg Arg Asn Asp			
215		190	195	200
216				
217	Asp Gly Asn Phe Pro Ala Val Gln Ile Pro Ser Ser Ser Thr Ser Ser			
218		205	210	215
219				
220	Pro Val Asn Gln Pro Thr Ser Thr Ser Thr Thr Ser Thr Ser Thr Thr			
221	220	225	230	235
222				
223	Ser Ser Pro Pro Val Gln Pro Thr Thr Pro Ser Gly Cys Thr Ala Glu			
224		240	245	250
225				
226	Arg Trp Ala Gln Cys Gly Gly Asn Gly Trp Ser Gly Cys Thr Thr Cys			
227		255	260	265
228				
229	Val Ala Gly Ser Thr Cys Thr Lys Ile Asn Asp Trp Tyr His Gln Cys			
230		270	275	280
231				
232	Leu			
233				
234				
235				

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1473 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: Fusarium oxysporum
  - (B) STRAIN: DSM 2672
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION: 97..1224
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

PAGE: 1

**SEQUENCE VERIFICATION REPORT**  
**PATENT APPLICATION US/09/189,028**

DATE: 08/25/1999  
TIME: 16:39:33

*INPUT SET: S33065.raw*

Line

Error

Original Text